Neurocomputing 188 (2016) 151–159

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Pedestrian identification based on fusion of multiple features and multiple classifiers

Xuan Wang [a,b], Qingxiang Wu [a,*], Xiaojin Lin [a], Zhiqiang Zhuo [a], Liuping Huang [a]

[a] *Key Laboratory of Optoelectronic Science and Technology for Medicine of Ministry of Education, College of Photonic and Electronic Engineering, Fujian Normal University, Fuzhou 350007, PR China*

[b] *College of Information Science and Engineering, Fujian University of Technology, Fuzhou 350007, PR China*

ARTICLE INFO

Article history:
Received 8 July 2014
Received in revised form
15 October 2014
Accepted 29 October 2014
Available online 14 December 2015

Keywords:
Texture features
Angle histograms
Color moments
Gabor filters
Feature fusion
Spiking neural networks

ABSTRACT

Finding a specific person from videos in surveillance systems is a challenging task. In the videos, different people cannot be the same in a whole body appearance. Based on this fact, this paper has proposed new methods based on fusion of textures, angle histograms and color moments to find a specific person. The human visual system can discriminate different objects quickly and efficiently. Inspired by on-center and off-center receptive fields in the visual system, a network model based on spiking neurons is proposed to extract texture features, and it has behaviors similar to Gabor filters. According to human body proportion, a person image is divided into three parts: head, torso and leg. Texture features of three parts are extracted by means of this network. Back propagation neural network, multi-class SVM and KNN are used as classifiers. For improving recognition rate, different fusion methods have been studied such as the fusion of texture features and other features in three body parts, and decision fusion using voting mechanism, probability combination etc. The experimental results for different methods are provided and the best fusion method is proposed. The technology of Compute Unified Device Architecture is applied in the experiments, which greatly reduces the running time for extraction of texture features.

1. Introduction

In video surveillance, it is pretty common to decide whether the object is pedestrian or not using its contour. In many cases, it is necessary to identify a specific person. This is usually achieved by the features of an individual iris, face, gait and fingerprint. However, when a person is far from the camera, it is impossible to extract individual iris features and fingerprint features, and the face is also vague. Similarly when a person does not walk in the direction faced to the camera, recognition rate based on face features will be greatly reduced. Therefore, color features and texture features for whole body of a person are used to improve pedestrian identification in the situation.

Color features and texture features of a person are distributed in whole body for a person. The distribution of color and texture of whole body cannot be the same for different persons because different persons have different appearances such as hair, face, skin, shape, and often wear different clothes. Therefore, a specific person can be discriminated by color features and texture features of the appearances.

Abdolahi used a method for gait recognition on the basis of texture features in [1]. Derbel compared texture descriptor with shape descriptor for pedestrian identification in [2]. Moctezuma detected persons by Histogram of Oriented Gradient (HOG) with Gabor filters in [3]. Daisy combined texture features and shape features to retrieve image in [4]. Using Gabor filters, Yussof in [5] detected interest points of object that are invariable even if the object rotates and its scale changes. Tao extracted color features and texture features to re-identify person in [6].

Single type of features always has some merits and drawbacks. Color can well reflect features of a person in color images, and it is difficult to reflect features of a person in an infrared image at night. Texture features are more robust to illumination change. Therefore, multiple types of features are fused to identify a specific person.

Besides color features and texture features, other features are also applied in human identification. Wang in [7] extracted silhouette-based features for human identification. Iwashita identified person in terms of gait and shadow biometrics in [8]. Zhao used the angles of edge pixels of human silhouette to construct angle histograms as gait features in [9]. Xu proposed a three-level framework to detect sudden pedestrian crossings in

* Corresponding author.
E-mail addresses: xwang197806@sina.com (X. Wang), qxwu@fjnu.edu.cn (Q. Wu).

http://dx.doi.org/10.1016/j.neucom.2014.10.114
0925-2312/

[10]. The framework extracted Local Binary Pattern (LBP) difference features, HOG features and Haar features of the targets. Huang improved biologically inspired model and applied it to object classification in [11].

The visual system has a powerful ability to extract keys information of different objects from complicated environment [12]. In videos, despite illumination change, shadow, occlusion and other interference factors, human always accurately identifies the target in a short time. This shows that the visual system not only has the function of memory storage, but also could extract key features of the target and remove unimportant and useless information. In recent decades, the visual nervous system has been attracting many scientists, and become a hot research domain. Spiking neuron model was first proposed by Hodgkin and Huxley in [13]. The model describes how the neurons exchange information encoded by spikes. As the complexity, in practice, a lot of simplified models have been derived. The most popular model is the conductance based integrate-and-fire model. Therefore, this neuron model is used to construct a spiking neural network (SNN) to extract texture features.

Ganglion cells, which are a type of neurons in the visual system, have on-center and off-center receptive fields. Based on the conductance based integrate-and-fire model and receptive field model, Wu applied RGB value of each image pixel to segment object in [14] and extracted the feature of object edge in [15]. Xie in [16] described how the process of taking advantage of parallel execution of Graphic Processing Units (GPUs) to segment color image. Wu applied hierarchical SNN to simulate visual attention in [17]. On-center receptive field and off-center receptive field can be both modeled by Gabor functions. Daugman proposed two-dimensional Gabor functions as spatial filters for image to simulate receptive fields in [18]. Fogel used Gabor functions for distinguishing texture in [19].

Receptive fields of spiking neurons can perform functions similar to Gabor filters. Rullen explored SNN simulating Gabor filters to detect human face in [20]. The spike train is encoded by the fire order of the neuron. Elmir extended Rullen's model to filter fingerprint images for fingerprint identification in [21]. The time consuming is the same problem for algorithms in Rullen's model and Elmir's model. If membrane potentials of neurons are calculated in serial programs, their algorithms will take a long time to extract the features.

In this paper a SNN model is proposed to extract texture features from gray images and play similar roles of Gabor filters. Texture features are used to classify different persons. The outputs are encoded by the fire rate of each output layer neuron of the network. Adjusting the firing threshold value of spiking neurons, noise and some unimportant information can be ignored. This function could not be achieved by Gabor filter. Experimental results show that the spiking neuron model could reach good effect. The process of feature extraction is slow if the dimension of image is high. In order to speed up the process of feature extraction, the technology of Compute Unified Device Architecture (CUDA) is also employed. In order to achieve high recognition rate, the texture features are fused with other multiple features such as angle histograms and color moments, and then multiple classifiers are trained and used to identify specific person. The main contribution of this paper is to propose parallel model inspired by the biological visual system, which can be used to extract texture features from a specific person and may help us to understand how the visual system extracts important information from images. A GPU based implementation scheme is proposed to solve time consuming problem of SNN. In addition, fusion methods of multiple features and multiple classifiers have been analyzed and used to improve recognition rate. The best fusion method has been proposed based on experimental results.

This paper is organized as follows: in Section 2 preprocessing of the datasets is introduced. Section 3 elaborates how to use SNN based on simplified conductance-based integrate-and-fire neurons to extract texture features and how angle histogram and color moment are constructed. Data fusion is also introduced. In Section 4, feature classification and fusion methods are described and experimental results are analyzed and compared. Section 5 concludes the paper and presents topics for further study.

2. Preprocessing pedestrian datasets

Experimental data are from the silhouettes and the color images from gait dataset A provided by CASIA [7]. The dataset contains walking images of 20 pedestrians. The walking images of 0 degree are selected for the test of algorithms.

Another group of experimental data is the Multi-Camera Videos (called MCV datasets) collected by Key Laboratory of Optoelectronic Science and Technology for Medicine of Ministry of Education, College of Photonic and Electronic Engineering Fujian Normal University. Gaussian mixture model in [22] is applied to extract the foreground of moving pedestrian. Because of the noise in the foreground, three-frame-difference method, whose threshold is set by Otsu, is used to determine moving region to detect walking people.

Image (a) in Fig. 1 is extracted by Gaussian mixture model. Image (b) is acquired by three-frame-difference. The rectangle in image (b) indicates moving region. Image (c) is a part of image (a) within moving region. Some images with extremely excessive noise and the images in which pedestrians block mutually are deleted to clean data.

There are five scenes with 39 pedestrians in the videos. Each scene is consisted of several video clips. Video clips are cataloged into single-person videos, or multi-person ones. All the pedestrians in the videos are labeled with names such as cbs, cj, cyf, dqj, lxj, lxj-2, zgr, zr, zr-2. The labeled persons can be used as a target person to be identified. If the same pedestrian wears different clothes, the label has same string name following a dash and a different number, for example, lxj and lxj-2 are the same pedestrian dressing differently. In the experiments the size of pedestrian image is normalized to 220×150. The pedestrian cartogram in five scenes is as follows:

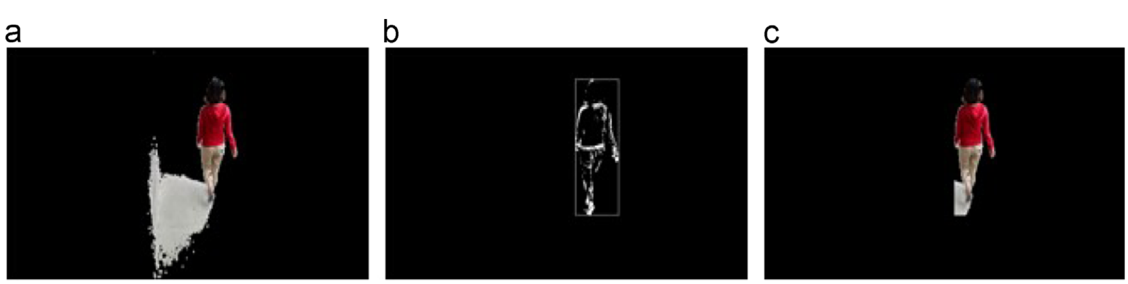

Fig. 1. Extracting the prospect of moving pedestrian.